DUT 37
Årgang 20, nr. 37, 2025

Boganmeldelser AI

Ruth Horak[1], Uddannelser & Studerende (US) Digital, Københavns Universitet

Abstract

AI – især NLP og store sprogmodeller – formår at engagere folk på tværs af discipliner og interesser; efterligningen af menneskelig sprogforståelse og interaktion giver ikke bare anledning til nysgerrig undren i forhold til teknologien, men sætter også tanker i gang om, hvad det "egentligt menneskelige" ved sprog, kommunikation, intelligens og tænkning er.

De tre bøger, jeg har valgt, har til fælles, at de ikke kun har et teknologifokus, men også kommer ind på de større, afledte spørgsmål: Melanie Mitchell henvender sig til "tænkende mennesker", Anders Søgaard forbinder et teknologisk med et filosofisk perspektiv, og Mark Coeckelbergh fokuserer på teknologifilosofiske og politisk-epistemologiske spørgsmål. Tilsammen giver min "bogpakke" en god introduktion til kunstig intelligens som felt og ridser nogle af de spørgsmål op, vi bør stille os selv og hinanden i de kommende år.

Melanie Mitchell (2019). Artificial Intelligence: A Guide for Thinking Humans. Pelican Books (419 sider).

Melanie Mitchells bog "Artificial Intelligence: A Guide for Thinking Humans" er en velskrevet introduktion til kunstig intelligens og har været min egen indgang til området. Bogen giver læseren en grundlæggende forståelse for, hvad AI er og hvilke metoder og tilgange der ligger til grund for de mest kendte anvendelser.

Melanie Mitchell, professor på Santa Fe Institute, har en PhD i datalogi og beskæftiger sig med kunstig intelligens, maskinlæring, kognitiv videnskab og komplekse systemer.

Som titlen indikerer, er bogen mere end en teknologifokuseret introduktion til kunstig intelligens. Mitchells reflekterende og kritiske betragtninger omkring teknologiens mere eksistentielle implikationer udspringer af solid, faglig ekspertviden, og hun vurderer potentialer og risici med afsæt i realisme og uden hype.

Bogen er fra 2019 og dermed skrevet før lanceringen af ChatGPT i november 2022. Det er dog ikke noget principielt problem, da bogen forholder sig til AI på det metodologiske og konceptuelle niveau, som ikke har ændret sig fundamentalt siden. Jeg ser det som bogens særlige styrke, at den giver lægfolk en grundlæggende, generisk forståelse for forskellige tilgange inden for kunstig intelligens. Til trods for de til tider komplekse og krævende beskrivelser af teknologien er bogen ret letlæst, og Mitchells kritiske refleksioner undervejs kan også engagere mere humanistisk orienterede læsere. Hvis man er interesseret i Mitchells syn på den nyeste udvikling og aktuelle temaer inden for AI, kan man følge hendes blog, som fint supplerer bogen.

Bogen består af fem overordnede dele med hver deres vinkel: baggrund, billedteknologi, spil og læring, sprogteknologi, forståelse og betydning. I bogens sidste kapitel, "Questions, answers and speculations", kommer Mitchell ind på fremtidsperspektiver og frekvente spørgsmål og bekymringer.

Første del giver et kronologisk overblik over AI-disciplinens historie og introducerer grundlæggende tilgange

[1] ruh@adm.ku.dk

og begreber (symbolsk vs. subsymbolsk, perceptroner, neurale netværk, maskinlæring m.m.). Allerede her ridser Mitchell de diskussioner op, som har fulgt udviklingen: spørgsmålet om, hvorvidt maskiner kan tænke, om Turing-testen som målestok, singularitetshypotesen og betydningen af menneskers kropslige, fysiske og emotionelle erfaring som forudsætning for læring og tænkning.

I bogens efterfølgende dele gennemgår Mitchell forskellige anvendelser af AI og beskriver de teknikker og tilgange, som har vist sig at være afgørende for feltets udvikling: kapitlet om behandlingen af visuelt input (computer vision) forklarer fx konvolutionelle neurale netværk (CNN) og deep learning samt forskellen mellem supervised og unsupervised læring. Her får man samtidig en forståelse for de problemer, dette kan medføre: fx overfitting og spørgsmålet om, hvad maskiner lærer, når de lærer. Maskiner lærer nemlig ikke altid det, vi mennesker antager, at de lærer, og kan desuden forstyrres igennem målrettede, ondsindede angreb bare ved at ændre bittesmå detaljer (adversarial attacks). På denne baggrund udfolder Mitchell de etiske problemer ved teknologien (bias, manglende pålidelighed og gennemskuelighed, indgreb i privatsfæren, databeskyttelse m.m.) og beskriver "the value alignment problem": hvordan sikrer vi, at de værdier, AI-systemer tilegner sig, stemmer overens med vores menneskelige værdier? Og: Kan vi mennesker blive enige om, hvad vores værdier er?

De næste kapitler i bogen handler om reinforcement learning, som sætter maskiner i stand til at lære at spille endda meget komplekse spil som fx Go. Men disse deep-learning-systemer møder udfordringer, når læringen skal generaliseres og overføres fra et område til et andet (transfer learning) – modsat mennesker, hvis intelligens netop består i at kunne overføre læring:

> *"For humans, a crucial part of intelligence is, rather than being able to learn any particular skill, being able to learn to think and to then apply our thinking flexibly to whatever situations or challenges we encounter. This is the true skill we want our children to learn […]" (Mitchell 2019: 217)*

Flere kapitler omhandler sprogteknologi, dvs. den AI-underdisciplin, der danner grundlaget for sprogmodel-baserede chatbotter som ChatGPT, Copilot, Gemini osv. Her gennemgår Mitchell afgørende teknikker og metoder som fx Recurrent Neural Networks (RNN), encoder og decoder, samt hvordan sprogets bestanddele omdannes til tokens og beregnes i form af vektorer. Bogen forklarer altså grundlæggende forudsætninger, men har ikke nået at få transformer-arkitekturen med, som i dag er den dominerende tilgang i moderne sprogmodeller. Her bliver man altså nødt til at supplere med yderligere læsning.

Bogens sidste del tager tråden op fra tidligere kapitler og kredser om mere overordnede spørgsmål om betydning og forståelse. Mitchell argumenterer for, at abstraktioner og mentale modeller er forankret i en konkret fysisk og kropslig erfaring, og at denne barriere er en væsentlig og hidtil uløst udfordring i udviklingen hen imod en potentiel "generel kunstig intelligens".

Det sidste kapitel handler om fremtidsudsigterne og spørgsmål som kreativitet, autonome supercomputere og hidtil uløste problemer. Det er velgørende, at Mitchell hverken er utopist eller dystopist; hun beskriver teknologiinhærente og afledte etiske udfordringer og advarer mod forhastet AI-autonomi, hype og antropomorfisering. Og ikke mindst opfordrer hun os til ikke at undervurdere vores egen menneskelige intelligens til trods for dens begrænsninger:

> *"What seems more likely to me is that these supposed limitations of humans are part and parcel of our general intelligence. The cognitive limitations forced upon us by having bodies that work in the world, along with the emotions and 'irrational' biases that evolved to allow us to function as a social group, and all the other qualities sometimes considered as cognitive 'shortcomings', are in fact precisely what enables us to be generally intelligent rather than narrow savants. I can't prove it, but I think it's likely that general*

intelligence can't be separated from all these apparent shortcomings, in humans or in machines." (Mitchell 2019, s. 365)

Det er netop denne kombination af dyb faglig viden, gode formidlingsevner og kritiske (meta-)betragtninger, der gør Mitchells bog værdifuld og oplysende i mine øjne.

Mark Coeckelbergh (2024). Why AI Undermines Democracy and What To Do About It. John Wiley & Sons (152 sider).

Jeg har valgt Mark Coeckelberghs tankevækkende bog, fordi den løfter problematikken omkring AI op på et samfunds- og teknologifilosofisk niveau uden at forfalde til et dystopisk og teknologideterministisk syn.

Mark Coeckelbergh er professor i medie- og teknologifilosofi ved Wiens Universitet. I sin forskning fokuserer han på etik og teknologi, især i forhold til robotteknologi og kunstig intelligens.

Coeckelberghs bog taler om AI i bred forstand: ikke kun algoritmer, men også data, infrastruktur, økonomi og økologi. Dette brede, holistiske syn ser jeg som en af bogens styrker, da hypen omkring generativ AI i øjeblikket flytter fokus fra de mange andre anvendelser af AI, som er lige så relevante, betydningsfulde og værd at reflektere kritisk over.

Efter min mening er bogen stærkest i første halvdel, hvor Coeckelbergh forholder sig kritisk til AI og taler om teknologiens ikke-neutralitet. Hovedbudskabet i Coeckelberghs bog er, at den måde, som AI udvikles og anvendes på i øjeblikket, underminerer grundlæggende principper og den vidensbase, som vores demokrati hviler på, og at AI ikke bidrager til det fælles bedste. Bag dette står grundantagelsen, at teknologi og AI ikke kun er værktøjer, men aktivt former vores tanker, handlinger og mål og dermed påvirker vores samfund. AI – og teknologi i almindelighed – er dermed selv politisk. Som en, der til dagligt arbejder med digital læringsteknologi, har jeg i mange år været optaget af dette aspekt: at teknologi aldrig er værdineutral og per se er udtryk for antagelser, holdninger og forventninger, og at vi skylder os selv og andre at være bevidste om dette, når vi interagerer med teknologi og udsætter andre for den.

Coeckelbergh argumenterer, at magten i forhold til AI i øjeblikket er asymmetrisk koncentreret hos kapitalistiske Big Tech-virksomheder og myndigheder, og at automatisering og bureaukratisk tankeløshed kan medføre alvorlige risici og trusler for borgerne. AI kan underminere demokratiske principper som frihed (i form af manipulation, overvågning, autoritære og antidemokratiske bevægelser), lighed (bl.a. i forhold til retfærdighed og fairness) og sammenhold (i form af polarisering, epistemiske bobler og ekkokamre).

Bogens mest spændende tanke for mig er denne: Et demokrati kræver særlige vidensmæssige (epistemiske) omgivelser. For at forme det vidensgrundlag, som aktiv og kritisk politisk deltagelse i et demokratisk samfund fordrer, skal vi som borgere udsættes for divergerende synspunkter, udfordres og møde friktion. Når AI-algoritmer lukker os ind i små, begrænsede vidensbobler og ekkokamre, gør de os samtidig sårbare over for mis- og desinformation, begrænser vores autonomi og danner grobund for mistillid og polarisering i samfundet, indbyrdes så vel som over for myndigheder, forskning og eksperter. Med andre ord: Når vores vidensmæssige handlekraft (epistemic agency) smuldrer, udhules samtidig vores politiske handlekraft (political agency) som oplyste, kritiske deltagere i et demokrati.

Bogens anden halvdel drøfter mulige løsninger. Coeckelbergh anbefaler bl.a. stærkere governance og reguleringer (også på globalt plan) samt en demokratisering af teknologiudviklingsprocessen gennem inkluderende participatoriske processer (democracy by design). Forslagene indbefatter en delvis eller hel afprivatisering af de teknologier og infrastrukturer, som danner grundlaget for AI, og ikke mindst et fokus på AI som kommunikationsteknologi, som aktivt kan bidrage til kommunikation og fællesskabsdannelse: ved at skabe

fælles videns- og erfaringsmæssige forudsætninger i stedet for epistemiske bobler og ekkokamre. Gang på gang fremhæver Coeckelbergh vigtigheden af kommunikation, her i den særlige betydning 'at gøre fælles, opbygge et fællesskab' (commun-ication), hvor man igennem fælles viden og erfaringer transcenderer sit eget ståsted. Tabet af epistemisk handlekraft og den medfølgende polarisering bør modarbejdes gennem en investering i uddannelse med et stærkt fokus på kritisk tænkning og diskussionsevner. Bogen afsluttes med fire anbefalinger til policy-makers: udligning af magtbalancen og indlejring af demokratiske værdier i teknologiudviking og innovation, styrkelse af eksisterende demokratiske institutioner kombineret med skabelsen af nye og mere participatoriske institutioner, investering i uddannelse, fokus på det fælles bedste gennem reguleringer, men også fællesskabsdannelse gennem forbedret kommunikation på tværs, fx understøttet af tekniske tiltag mod ekkokamre og polarisering og et uddannelsessystem, der fremmer kritisk tænkning.

Denne del finder jeg mindre overbevisende, om end jeg sympatiserer med forslagene. Dels kan man sætte spørgsmålstegn ved, hvor realistiske forslagene er (selv frygter jeg, at nogle af dem desværre tangerer det utopiske), dels ligger realiseringen uden for den individuelle læsers umiddelbare indflydelse og kan derfor efterlade læseren med en følelse af magtesløshed: For hvis løsningen er afhængig af flertallet (demokrati) og policy-makers, hvor stor en rolle spiller jeg så egentlig selv? Omvendt kan man indvende, at det netop er her, demokratiske processer og vores egen epistemiske handlekraft skal stå sin prøve, og at denne proces starter med den enkelte.

Mest værdifuld for mig var Coeckelberghs understregning af teknologiers politiske ikke-neutralitet (teknologier er altid også politiske) og erkendelsen af den kausale sammenhæng mellem epistemic agency og political agency. Måske er det i virkeligheden her, processen starter: med bevidstheden om sammenhængen mellem fællesskab, epistemisk handlekraft og oplyst politisk deltagelse, og hvordan AI kan påvirke denne sammenhæng.

Anders Søgaard (2022). Kunstig intelligens bagfra. Aarhus Universitetsforlag (96 sider).

Anders Søgaards bog er valgt af to årsager: Dels for at inkludere en dansk publikation, dels fordi forfatteren både er professor i sprogteknologi på KU, lingvist, forfatter og filosof, og bogen derfor hele tiden bevæger sig frem og tilbage mellem en informativ, saglig fremstilling af faktuel viden om AI, lingvistiske betragtninger og et filosofisk metaniveau, hvor Søgaard reflekterer over teknologien i forhold til aspekter som bevidsthed, forståelse, fri vilje og subjektivitet. Samtidig får man indblik i Søgaards egen tilgang til sprogteknologi og de erkendelser og spørgsmål, der har optaget ham gennem tiden.

Bogen er inddelt i fem kapitler, som beskæftiger sig med disciplinens historie, maskiners læring, menneskelig læring, forklarlighed og transparens og slutteligt refleksioner over kunstig intelligens som et spejl for os mennesker.

Ligesom Mitchell giver bogen et historisk overblik over udviklingen af kunstig intelligens som disciplin. Søgaards bog er dog en del kortere og går derfor mindre i detaljer. Til gengæld fokuserer den især på sproglige aspekter og NLP – til stor glæde for en humanist med sproglig baggrund som mig.

I kapitlet om maskinlæring beskriver Søgaard de udfordringer, som sproglig variation udgør for kunstig intelligens, hvordan sprogmodeller lærer om sprog og hvordan genbrug i form af prætrænede modeller kan reducere omkostninger, energiforbrug og klimaaftryk, men også nedarve bias, fejl og mangler. Her er der et par sider, der er forholdsvis teknisk orienterede, men Søgaard er god til at forklare og oversætte tekniske mekanismer til almindeligt menneskesprog.

Den menneskelige hjerne og menneskers læring danner udgangspunktet for at forklare den historiske udvikling fra perceptroner til neurale netværk og begreber som back-propagation og reinforcement learning.

I kapitlet om forklarlighed kredser Søgaard om et grundlæggende problem ved forklaringer: Hvad skal der til for at leve op til kravet om forklarlighed? Hvad er en god forklaring? Hvordan rammer man balancen mellem det overfladiske og det alt for detaljerede, og hvor gode – eller omvendt: fejlbehæftede og misvisende – er menneskelige forklaringer i grunden?

Det sidste kapitel handler om det, som Søgaard selv betegner som den røde tråd i sit arbejde med AI: AI som et spejl, vi kan holde op foran os selv og vores samfund og som kan udfordre den forståelse, vi har af os selv. Teknologien er anledning til at se kritisk på os selv og grave dybere i begreber som fri vilje, subjektivitet og bevidsthed. De sidste sider handler om bekymringer, især Søgaards største bekymring: at AI kan skabe større ulighed ved at begunstige nogle befolkning(sgrupp)er på bekostning af andre. Ligesom Coeckelbergh understreger Søgaard, at AI har potentialet til at gøre begge dele: forstørre uligheden eller mindske den, men at denne udvikling ikke er skæbnebestemt – den kræver, at vi tager et valg.

For mig ligger bogens særlige styrke i, at den er rig på filosofiske refleksioner, undrende spørgsmål og beskrivelser af videnskabelige eksperimenter og erkendelser med mennesker som udgangspunkt; en kombination, der sætter tanker i gang om kunstig vs. menneskelig intelligens, meningsskabelse, fortolkning, sproglig forståelse og bevidsthed.

En af øjenåbnerne for mig var forfatterens (med reference til filosoffen Daniel Dennetts) skelnen mellem sprogforståelse og bevidsthed om sprogforståelse. Computere kan have sproglig forståelse og anvende sprog meningsfuldt, men de er ikke bevidste om deres egen forståelse – hvilket er, hvad der adskiller dem fra os mennesker.

Hvis man vil læse om kunstig intelligens (og især sprogteknologi) på dansk og sætter pris på kritisk metarefleksion, kan bogen anbefales på det varmeste.

Referencer

Mark Coeckelbergh (2024). *Why AI Undermines Democracy and What To Do About It*. John Wiley & Sons. 152 sider.

Melanie Mitchell (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Pelican Books. 419 sider.

Anders Søgaard (2022). *Kunstig intelligens bagfra*. Aarhus Universitetsforlag. 96 sider.